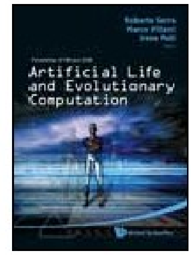


ARTIFICIAL LIFE AND EVOLUTIONARY COMPUTATION

Proceedings of Wivace 2008

Venice, Italy, 8 – 10 September 2008

edited by **Roberto Serra** *(University of Modena and Reggio Emilia, Italy)*, **Marco Villani** *(University of Modena and Reggio Emilia, Italy)*, & **Irene Poli** *(Ca' Foscari University, Italy)*

The Italian community in Artificial Life and Evolutionary computation has grown remarkably in recent years, and this book is the first broad collection of its major interests and achievements (including contributions from foreign countries). The contributions in Artificial Life as well as in Evolutionary Computation allow one to see the deep connections between the two fields. The topics addressed are extremely relevant for present day research in Artificial Life and in Evolutionary Computation, which include important contributions from very well-known researchers. The volume provides a very broad picture of the Italian activities in this field.

Table of Contents

Readership: Graduate students, academics and researchers in the field of complex systems, artificial intelligence and robotics.

340pp
Pub. date: Nov 2009
eISBN: 9789814287456

Purchase book at WorldSciBook

"ADMISSIBLE METHOD FOR IMPROVED GENETIC SEARCH IN CELLULAR AUTOMATA MODEL (AMMISCA)": A STRATEGY IN GENETIC CALIBRATION - PRELIMINARY RESULTS

R. UMETON* and S. DI GREGORIO

Department of Mathematics, University of Calabria
Arcavacata di Rende, CS 87036, Italy.
www.unical.it

Genetic Algorithms (GAs) are widely used to incrementally reach admissible solutions for hard problems such as parameter tuning in Cellular Automata (CA) models. This paper presents a genetic strategy, specifically developed for CA model calibration, exploiting the circumstance that the considered CA parameters have a physical meaning. The proposed approach has proved to be comparable and, in some cases outperforming, if compared with the standard GA proposed by Holland. As a further result, the goodness of the proposed genetic strategy opens the door to genetic tuning algorithms lacking of a standard crossover operator. Results, though preliminary, can be considered encouraging and routes to a wider analysis of the proposed approach.

Keywords: genetic algorithms; model calibration; cellular automata; lava flows.

1. Introduction

In the field of risk assessment and hazard mitigation, event simulation and predictor models have acquired a relevant position. In fact, through simulation of reliable models, risks associated to such processes can be evaluated and possibly predicted and contrastated.

Cellular Automata[1] (CA) proved[2–5] to be a valid choice in simulating natural phenomena such as landslides, erosion processes, lava and pyroclastic flows. They are parallel computing models, discrete in space and time, whose dynamics is determined by the application of local rules of evolution defining the CA transition function. In particular, above cited examples

*Corresponding author. Tel.: +39 0984 496467; fax: +39 0984 496410.
Email addresses: `umeton@mat.unical.it` (R. Umeton), `dig@unical.it` (S. Di Gregorio).

are based on the Di Gregorio and Serra's approach[6] for the modelling of spatially extended dynamical systems. Models based on this approach generally depend on many parameters, which must be provided with the highest possible accuracy in order to obtain satisfactory results in simulating the considered phenomenon. To do this, a parameter tuning phase through standard GA has been successfully applied in previous works[7–10] .

Genetic Algorithms (GAs)[11,12] are parallel, general-purpose, search algorithms inspired by Genetics and Natural Selection. They simulate the evolution of a population of candidate solutions of a specific search problem by favouring the "survival" and the "recombination" of the best ones, in order to obtain better and better solutions. These family of algorithms have acquired an important role in all those fields dealing with intrinsically-hard problem lacking of dedicated heuristics or ad-hoc algorithms.

This paper proposes the definition of AMMISCA, a genetic strategy, and its application to the parameter tuning of the SCIARA-R7[13] CA model for lava flow simulation and forecasting. The next section presents the SCIARA-R7 simulation model. Section 3 details the AMMISCA genetic strategy, while the fourth section discusses obtained results. Conclusions are reported at the end of the paper.

2. The SCIARA-R7 model

The physical behaviour of lava flows can be partially described in terms of Navier-Stokes equations. Analytical solutions of these differential equations are a hopeless challenge, except for few simple, not realistic, cases. The complexity of the problem resides both in the difficulty of managing irregular ground topography and in complications of the equations, that must also be able to account for flows, exhibiting a wide diversity in their fluid-dynamical behaviour due to cooling processes. An alternative approach to PDE numerical methods for Navier-Stokes[14] (or more complex) equations is offered by Cellular Automata (CA). As announced above, they are computational models assuming discrete space/time and easily implementable on parallel computers. CA SCIARA-R7 for lava flows is derived from SCIARA[4] where the space is a plane, divided in hexagonal cells; each cell is characterised by a state, that specifies the mean values of physical quantities in the cell (e.g. substate altitude) and embodies a computing unit. This unit updates synchronously the substate values according to a transition function on the basis of substate values of the cell and its adjacent ones. The transition function is applied by the sequential computation of "elementary processes", that account for the phenomenon features.

From a formal point of view SCIARA-R7 is stated by the septuple $SCIARA\text{-}R7 = \langle R, L, X, S, P, \sigma, \gamma \rangle$, where

- $R = \{(x, y) | x, y \in \mathbb{N}, 0 < x < l_x, 0 < y < l_y\}$ is the set of identical hexagonal cells identified by integer co-ordinates in the finite region where the phenomenon evolves.
- $L \in R$ specifies the lava source cells (i.e. craters).
- X identifies the geometrical pattern of cells that influence the cell state change. They are, respectively, the cell itself and its adjacent cells: $X = \{(0, 0), (0, 1), (0, -1), (1, 0), (-1, 0), (-1, 1), (1, -1)\}$.
- $S = Q_A \times Q_{th} \times Q_T \times {Q_O}^6$ is the set of states; more in detail, Q_A is the altitude of the cell, Q_{th} is the thickness of lava inside the cell, Q_T is the lava temperature and ${Q_O}^6$ rappresent lava outflows (6) from the central cell towards the adjacent ones.
- $P = \{p_{clock}, p_{TV}, p_{TS}, p_{chlV}, p_{chlS}, p_{adher}, p_{cool}\}$ is the set of global parameters, in which:
  - p_{clock} is the time corresponding to a CA step
  - p_{TV} is the lava temperature at vent
  - p_{TS} is the lava solidification temperature
  - p_{chlV} is the characteristic length at the vent temperature
  - p_{chlS} is the characteristic length at the solidification temperature
  - p_{adher} is the constant adherence of lava passing on a cell
  - p_{cool} is the cooling parameter
- $\sigma : Q^{6+1} \rightarrow Q$ is the deterministic state transition function, which is simultaneously applied to all cells of the CA.
- $\gamma : Q_{th} \rightarrow \mathbb{N} \times Q_{th}$ specifies the emitted lava from source cells at the CA step $t \in \mathbb{N}$.

In order to evaluate the goodness of simulations obtained with the detailed model, we have adopted the evaluation function $e_2 = \sqrt{\frac{R \cap S}{R \cup S}}$ where R and S represent the area covered by simulated and real lava flow, respectively; this evaluation function is then used to compute the fitness associated to each simulation in the genetic process.

The next Section describes the proposed genetic strategy for optimizing SCIARA-R7 parameters by means of the AMMISCA approach.

3. AMMISCA in detail

AMMISCA, the acronym of AdMissible Method for Improved genetic Search in Cellular Automata, is a genetic strategy exploiting the circumstance that each element of the set of parameter to be tuned (P) has a physical meaning. For instance, if $parent_A$ and $parent_B$ expresses the p_{chlV} SCIARA-R7 parameter (which represents a "threshold" for lava mobility) with values 15 and 25 meters respectively, it can be erroneous to assign the next offspring to an improbable value of 50 meters (which is too *distant* from the parent contributes). As anticipated above, the main, characterizing, difference between a standard Holland GA and AMMISCA regards the field which they have been designed for (Cfr. Fig. 1). While the standard GA is a general purpose optimizer, the second one has been designed for the resolution of those problems in which parameters encoded in the individual have a physical correspondence. When this physical correspondence exists, our algorithm takes advantage of it, thanks to the different crossover strategy implemented, which strictly preserves previous obtained results.

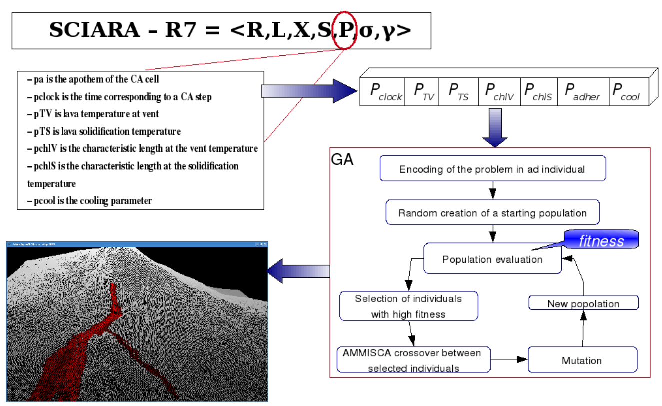
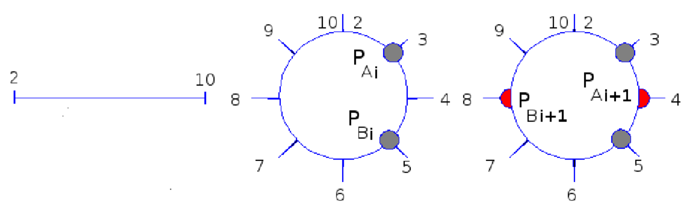


Fig. 1. The tuning process: (1) select the part of the model that has to be tuned: the parameter set in our case; (2) encode this parameter set in the individual; (3) run the Genetic Algorithm in order to let admissible solutions evolve and recombine, favouring better solutions: in our case the fitness is evaluated through the function e_2; (4) extract the parameter set that gave the most realistic simulation; (5) adopt this set to complete the lava forecasting model.

The basic idea within the AMMISCA strategy is to go beyond the preservation of promising schemes through a different crossover, based on arithmetic average: while a one-point crossover ($ONEPT$), using a randomly selected crosspoint, can transform parent strings (e.g. AAAAA, BBBBB) into quite different strings (e.g. AABBB, BBAAA), our crossover calculates for each parameter the average value between parent ones (as proposed in Linear crossover[15] method with $weight = 0.5$), and assigns it to the next generation allele. From two parents we get only one offspring; moreover, this single individual might be too much specialized and the average-driven recombination seems to converge too much rapidly. In order to solve these problems, a sort of *anti-dimidium* is introduced. In AMMISCA, as in standard GAs, we have a range for each parameter encoded in the individual, and two points inside the range rappresenting the value of the parameter introduced by parents. If we shift from a linear range to a closed one (Cfr. Fig. 2), we obtain a circumference where minimum and maximum of the range coincide and, while the average value is assigned to the first offspring (i.e. $P_{A_{i+1}} = (P_{A_i} + P_{B_i})/2$) as the logical middle-point between parent values, the *anti-dimidium* is calculated as the point diametrically opposite to it (i.e. $P_{B_{i+1}} = P_{A_{i+1}} + (P_{max} + P_{min})/2$).



(a) Linear range of a parameter p.

(b) Closed range with values of p exhibited by two individuals.

(c) Closed range with values of p assigned to the offspring off individuals in (b) according to AMMISCA "average version".

Fig. 2. The shift from the linear range to the closed one along with average and *anti-dimidium* definition.

An idea, subtended by the introduction of *anti-dimidium*, concerns the following problem: some couples of parameters in SCIARA are "antagonist". This means that similar results could be obtained increasing the value of the former parameter and decreasing the value of the latter one. Hence, different clusters of good values of parameters may exist. Then, AMMISCA always proposes an "internal" ($P_{A_{i+1}}$ in Fig. 2) allele and an "external" ($P_{B_{i+1}}$) one: the former searches for a specialization of parents, while the latter explores values out of the interval defined by parents. Finally, in the context of SCIARA clustered parameters, AMMISCA conveniently derives a new offspring by composing "internal" and "external" alleles (Cfr. last line of following pseudo-code block). Besides the application of the mere average for model calibration as described above, this work presents two further variants of the algorithm which consider different offspring calculations. In particular, the first version uses the fitness associated to every parent in order to weigh their contribution and thus is labeled as a "fitness weighted average" ($FWAVG$); indeed, the more a parent is promising, the *closer* the allele will be to it. The second variant chooses a random point inside the sub-interval delimited by parents (denoted as $RWAVG$ and proposed in Heuristic crossover[16]). The pure application of the versions detailed above could result too fitness-driven and interfer with crossover function and research space inspection (the first variant, fitness weighted average), or could take longer to solve easy problems (e.g. a maximum values search in a simple cusp by means of the second variant, randomly weighted average). Then, the combination of internal ad external alleles permits to embank this problem. In order to fix details given up to now, a pseudo-code-block that states the AMMISCA crossover is now proposed.

BEGIN: AMMISCA_crossover_function()
{
  crossmode = get requested crossover type *//one in {ONEPT, AVG, FWAVG, RWAVG}*
  for each (parameter p **in** P encoded in the individual)
    P_A = value of parameter p expressed by $parent_A$. Same for P_B.
    $P_A{}'$ = value of parameter p that will be assigned to $offspring_A$. Same for $P_B{}'$.
    $range_{min}$ and $range_{max}$ are minimum and maximum value assignable to parameter p
    if (crossmode==ONEPT) applyStandardCrossoverByHolland(P_A,P_B,$P_A{}'$,$P_B{}'$);
    else if (crossmode==AVG) $P_A{}' = (P_A + P_B)/2$;
    else if (crossmode==FWAVG)
      $P_A{}' = (P_A * fitness_A + P_B * fitness_B)/(fitness_A + fitness_B)$;
    else if (crossmode==RWAVG) *//uses only positive random numbers*
      $P_A{}' = (P_A * random_1 + P_B * random_2)/(random_1 + random_2)$;
    if ($P_A{}' == (range_{min} + range_{max})/2$)
      $P_B{}'$ = choose randomly, with same probability, between $range_{min}$ and $range_{min}$
    else $P_B{}' = P_A{}' +$ half round of the range; *//anti-dimidium*
    if ($P_B{}' > range_{max}$) $P_B{}' = P_B{}' - range_{max} + range_{min}$
    if (crossmode $\neq ONEPT$) swap $P_A{}'$ and $P_B{}'$ with proability 0.5; *//alleles composition*
} **END**;

In the next section we briefly present the main results achieved by AMMISCA applied to the calibration of the model SCIARA-R7.

4. AMMISCA Results

In order to validate the genetic strategy for a parameter tuning task, AMMISCA is used for the calibration of SCIARA-R7 model applied to the Nicolosi lava flow event which occurred at Mt Etna (Italy) in 2001.

We consider different classes of tests: first, we use many seeds and few GA generations (50 seeds and 10 generations, Cfr. Fig. 3 for setup details), and subsequently we adopt the most promising seeds for further GA generation computation (i.e. the most promising seed for 100 generations). In Table 1, the first test as a result of about 21000 individual evaluations is presented, where the four adopted algorithms (single-point crossover and AMMISCA in its three versions) are compared by means of the contribution of the best average in the individual pool and the best individual. The most promising seed of each algorithm is further executed for ninety more generations and Fig. 4 displays both fitness trend and time comparison.

Numbers of seeds	50

GA Setup	
Parameters num.	7
Initial number of individuals	16
Individuals replaced at each step	8
Crossover probability	1
Mutation probability	0,083
GA steps	10

Individual composition			
Parameter *(unit)*	NO. of bits	Range-min	Range-max
clock *(s)*	8	60	240
TV *(K)*	0	1323	1323
TS *(K)*	8	1023	1173
chlV *(m)*	4	0,1	5
chlS *(m)*	4	5,1	10
adher *(m)*	8	0,05	10
cool *(m^3/K^3)*	8	10^{-19}	10^{-12}

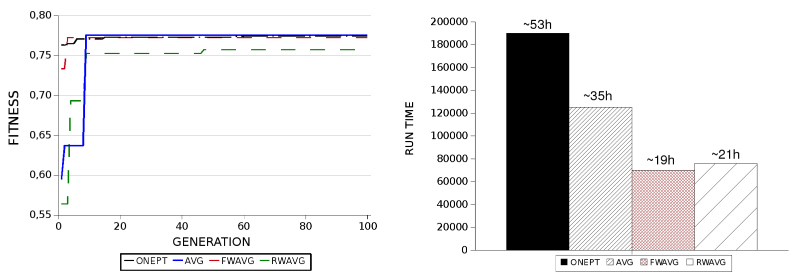
Fig. 3. The GA setup for the first class of test.

As a result, the AMMISCA strategy proves to be valid and promising, being able to outperform standard GA with single-point crossover, both in terms of obtained fitness and execution times. Table 1 and Fig. 4 indicate that AMMISCA obtains the best individual in both test cases (10 and 100

Table 1. First test set: generation-by-generation, for 10 generations, which algorithm gives the best results, over 50 seeds evaluation for each algorithm.

Generation	Best average fitness	Best individual
1	AVG	ONEPT
2	AVG	ONEPT
3	AVG	FWAVG
4	ONEPT	FWAVG
5	ONEPT	FWAVG
6	ONEPT	FWAVG
7	ONEPT	FWAVG
8	ONEPT	AVG
9	ONEPT	AVG
10	**ONEPT**	**AVG**

Note: ONEPT is one-point crossover; AVG is AMMISCA average version; FWAVG is fitness weighted average version; RWAVG is randomly weighted average version.



(a) Fitness trend evolution of each algorithm over 100 GA generations as emerged in second test set: keep running the most promising (at 10th generation) seed for each algorithm for 90 generations more.

(b) Total time required by each algorithm over the two test sets, remarking the fact that some of them explored search zones ignored by others, at least with respect to p_{clock} parameter.

Fig. 4. Second test set results and global time required by each algorithm to complete the two tests.

GA iterations), giving thus the most precise lava event simulation. Besides these results, AMMISCA chooses a set of individuals characterized by a high P_{clock} value leading to faster computations and lower execution times;

such P_{clock} were not taken into account by the Holland search strategy.

5. Conclusions and future developments

Results can certainly be considered encouraging for this initial stage in the study of AMMISCA genetic strategy. Moreover, besides the fact that AMMISCA gives rise to the most precise lava simulation, it is interesting to note that we achieve the best solution (in terms of fitness and required time) without a standard crossover phase as defined by Holland. Furthermore, these results route to ad-hoc tuning techniques for CA models that are similar to the analyzed one, that are CA models where the parameter set has a physical meaning.

As a consequence of these preliminary results, AMMISCA can be more deeply inspected as an alternative to a standard GA algorithm. The game plan for future work is to study the AMMISCA conduct in the calibration of SCIARA model for factitious lava events[17] (the best simulation is considered as the real lava event). In fact, we can better compare standard GAs and our family of algorithms with respect to an artificial lava event so that theoretically the global optimum can be achieved during calibration. To be more precise, the referring artificial simulation can be either the simulated lava event obtained with Holland's GA or the one obtained with AMMISCA "average version" (respectively the first and the second simulation whose fitness is rappresented in Fig. 4). Subsequently, the second step in this validation plan is to use our family of algorithms to calibrate other macroscopic CA models, tuned with standard GA in the past, such as SCIDDICA[2] , PYR[5] and SCAVATU[3] . Eventually, through the analysis of AMMISCA behaviour on cited models, we can derive a study of fitness landscape[17] and reach a more accurate idea of the AMMISCA convergence process.

Acknowledgments

Authors thank Dott.ssa Maria Vittoria Avolio, Dott. William Spataro, Dott. Rongo Rocco and Dott. Donato D'Ambrosio for their support during the development of the work and for their advices in the paper redaction.

References

1. J. von Neumann, *Theory of Self Reproducing Automata* (University of Illinois Press, 1966).

2. S. Di Gregorio, R. Rongo, C. Siciliano, M. Sorriso-Valvo and W. Spataro, *Phys. Chem. Earth Pt. A* **24**, 97 (1999).
3. D. D'Ambrosio, S. Di Gregorio, S. Gabriele and R. Gaudio, *Phys. Chem. Earth Pt. B* **26**, 33 (2001).
4. G. M. Crisci, S. Di Gregorio, R. Rongo and W. Spataro, *J. Vol. Geo. Res.* **132**, 253 (2004).
5. G. M. Crisci, S. Di Gregorio, R. Rongo and W. Spataro, *Fgcs* **21**, 1019 (2005).
6. S. Di Gregorio and R. Serra, *Fgcs* **16**, 259 (1999).
7. D. D'Ambrosio, S. Di Gregorio and G. Iovine, *Nhess* **3**, 545 (2003).
8. P. M. Atkinson, G. M. Foody, S. Darby and F. Wu, *Brains versus Brawn - Comparative Strategies for the Calibration of a Cellular Automata-based urban growth model*, in *GeoDynamics*, (CRC Press, 2004).
9. B. Straatman, R. White and G. Engelen, *Comput. Environ. Urban* **28**, 149 (2004).
10. D. D'Ambrosio and W. Spataro, *Parallel Comp.* **33**, 186 (2007).
11. J. H. Holland, *Nonlinear environments permitting effcient adaptation*, in *Computer and Information Sciences II*, (New York: Academic, 1967).
12. J. H. Holland, *Adaptation in Natural and Articial Systems* (University of Michigan Press, Ann Arbor, 1975).
13. M. Avolio, G. M. Crisci, S. Di Gregorio, R. Rongo and R. Umeton, Introduction of more physical features in the cellular automata model for lava flows sciara: preliminary results regarding the viscosity (2007), Asia and Oceania Geosciences Society.
14. H. Lamb, *Hydrodynamics* (Cambridge University Press, 1879).
15. A. H. Wright, Genetic algorithms for real parameter optimization, in *Foundations of genetic algorithms*, ed. G. J. Rawlins (Morgan Kaufmann, San Mateo, CA, 1991) pp. 205–218.
16. A. H. Wright, Genetic algorithms for real parameter optimization, in *Foundations of Genetic Algorithms, First Workshop on the Foundations of Genetic Algorithms and Classier Systems*, ed. G. J. Rawlins (Morgan Kaufmann, San Mateo, CA, 1990) pp. 205–218.
17. D. D'Ambrosio, W. Spataro and G. Iovine, *Comput. Geosci.* **32**, 861 (2006).